Econometrica, Vol. 81, No. 1 (January, 2013), 153–171

THE REVEALED PREFERENCE THEORY OF STABLE AND EXTREMAL STABLE MATCHINGS

BY FEDERICO ECHENIQUE, SANGMOK LEE, MATTHEW SHUM, AND M. BUMIN YENMEZ[1]

We investigate the testable implications of the theory of stable matchings. We provide a characterization of the matchings that are rationalizable as stable matchings when agents' preferences are unobserved. The characterization is a simple nonparametric test for stability, in the tradition of revealed preference tests. We also characterize the observed stable matchings when monetary transfers are allowed and the stable matchings that are best for one side of the market: extremal stable matchings. We find that the theory of extremal stable matchings is observationally equivalent to requiring that there be a unique stable matching or that the matching be consistent with unrestricted monetary transfers.

KEYWORDS: Revealed preference theory, two-sided matching markets, stability, extremal stability, assignment game.

1. INTRODUCTION

THIS PAPER STUDIES the testable implications of *stability* in two-sided matching markets. In the spirit of classical revealed preference analysis, we suppose that one can observe matchings, but not agents' preferences, and we want to understand the empirical content of matching theory.

We give simple conditions that characterize the observable restrictions of the theories of stable matching with transfers and without transfers, and the stable matchings that are optimal for each side of the market (extremal stable matchings). The model with transfers turns out to be strictly more restrictive than the model without transfers and is observationally equivalent to extremal stable matchings.

The revealed preference problem in matching presents unique challenges. In classical revealed preference theory, if Catherine chooses option A over option B, then we may infer that she prefers A over B. In a two-sided model, the situation is much more complicated: If Catherine matches with Jules and not with Jim, it may be because she likes Jules best, but it may also be because Jim is matched to someone *he* prefers over Catherine. Jim's preferences are, however, as unobservable as Catherine's. Hence, matching data do not unambiguously resolve the direction of revealed preferences. This problem is a crucial challenge: it is intrinsic to two-sided models and most empirical studies of matching circumvent the problem by assuming unlimited transfers among

[1]This paper evolved from Echenique, Lee, and Yenmez (2010) and contains generalizations of the theoretical results in Echenique, Lee, and Shum (2010), both of which are obsolete now. We thank Lars Ehlers for questions that motivated some of the current research. We are very grateful to the editor and anonymous referees for their comments on a previous draft.

 DOI: 10.3982/ECTA10011

the agents.[2] Ours is the first paper to provide a complete revealed preference characterization of stable matchings.

The literature on stable matching has grown rapidly, delivering beautiful theoretical results and important empirical applications. It is, however, largely a normative theory. The applications of matching theory deal with how markets should be designed (Roth (2008), Sönmez and Ünver (2011)).[3] As a positive theory, stable matching is not well understood.

For example, much is known about the two canonical models of matching: the model with no monetary transfers (the Gale and Shapley (1962) college-admissions model) and the model with transfers (the Shapley and Shubik (1971) assignment game), but we do not know how to *empirically* distinguish one model from the other. The marriage market is a popular application of both models, but one cannot observe transfers and thus understand which model is more appropriate.[4] We want to know which model is more appropriate when we only observe who matches with whom, but no preferences or transfers are observed.

Another example relates to extremal stable matchings, the standard outcomes in matching market design. In the absence of transfers, there are two distinguished stable matchings: each is the best stable matching for one side of the market and the worst for the other side. Centralized market clearinghouses, such as the National Resident Matching Program in the United States or the recent designs of public school choice programs, implement an extremal stable matching, but would a decentralized market select a similar outcome? To answer this question, we need to understand when observed matchings are compatible with the theory of extremal stable matchings.

We focus on a general notion of data, which we call aggregate matchings. In an aggregate matching, individuals on each side of the market are summed up into cells on the basis of their observed characteristics, such as age, educational attainment, or employment sector. Empirical researchers studying marriage, for instance, typically use aggregate matchings (Choo and Siow (2006)).

We assume that all individuals with the same characteristics are identical and have identical preferences. This is a strong assumption, but without it, the theory has no testable implications: once one allows for enough heterogeneity in preferences among observationally identical individuals, any matching could be trivially rationalizable. Thus, we focus on the restrictive case of no preference heterogeneity among individuals with the same characteristics. This

[2]Under this assumption, they can focus on the matchings that maximize social surplus.

[3]The existing literature often notes that actual markets use stable matching mechanisms (Roth (2002)), which is a positive finding. The thrust of the literature is, however, normative.

[4]In some empirical settings, such as labor markets, it may be possible to observe transfers. In a number of empirical applications of matching models, however, transfers are not observed; these include marriage markets (Becker (1973), Choo and Siow (2006)), auto-parts markets (Fox (2008)), and venture capital markets (Sørensen (2007)).

can be considered a "worst case" under which rationalizability is still possible.[5] Different assumptions on what may be observed (e.g., if one could partially observe preferences or some form of transfers) and on agent heterogeneity will affect our conclusions. One message of our paper is that to distinguish the theories that we find observationally equivalent, richer data sets are required.

2. MODEL

We present population versions of the two standard matching models: the model with nontransferable utility (NTU) and the model with transferable utility (TU). Our notion of population matching is termed *aggregate matching* and is distinguished from the more familiar notion of "individual" matchings.

The models feature two (disjoint) finite sets, M and W. The set M is a set of *types of men*, while W is a set of *types of women*. We enumerate M as $\{m_1, \ldots, m_{|M|}\}$ and W as $\{w_1, \ldots, w_{|W|}\}$. We are given a list $K = (K_i)_{i \in M \cup W}$ of nonnegative real numbers; K_m is the number (or the mass) of men of type m and K_w is the number of women of type w. A *matching* is an $|M| \times |W|$ matrix $X = (x_{m,w})$ such that $x_{m,w} \in \mathbf{R}_+$, $\sum_w x_{m,w} \leq K_m$ for all m, and $\sum_m x_{m,w} \leq K_w$ for all w. The number $x_{m,w}$ is the number (or the mass, or the probability) of men of type m matched to women of type w. We assume that there are no "singles" or unmatched agents; that is, $\sum_m K_m = \sum_w K_w$, and $\sum_w x_{m,w} = K_m$ for all m, and $\sum_m x_{m,w} = K_w$ for all w.[6]

We allow for noninteger values of $x_{m,w}$ to accommodate random matchings in our framework. For instance, Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005) and Kesten and Ünver (2009) studied the probabilistic assignment of school seats to students, and one may ask if a randomized matching such as this is consistent with stability for some preferences of the students and priorities of the schools. Our results are applicable to these randomized matchings as well. (See Echenique, Lee, and Yenmez (2010) for a complete discussion of this issue.)

The standard model of individual matchings results when we have $K_i = 1$ for $i \in M \cup W$ and $x_{m,w} \in \{0, 1\}$.

[5]This differs from the approach in existing empirical applications of matching theory, which assumes transferable utilities (see, e.g., Choo and Siow (2006), Fox (2008), or Galichon and Salanie (2009)), but allows for heterogeneous preferences at the individual level. Echenique (2008) and Chambers and Echenique (2009) studied the revealed preference problem for a collection of individual-level matchings, which differs from the general notion of matching data considered in the present paper.

[6]We rule out singles for expositional simplicity, but the results are easily extended. In most actual marriage data, we only observe formed couples, so assuming that there are no singles is not a problem for most empirical applications to marriage. As a result, we do not define individual rationality because it has no empirical bite when there are no singles. It is straightforward to modify our setup to study individual rationality (Echenique, Lee, and Yenmez (2010)).

2.1. *Matching With Nontransferable Utility*

The primitives of the model are represented by a four-tuple $\langle M, W, P, K\rangle$, where M, W, and K are as described above, and P is a *preference profile*—a list of preferences P_m for every type of man m and P_w for every type of woman w. Each P_m is a linear order over W and each P_w is a linear order over M. The weak order associated with P_m (P_w) is denoted by R_m (R_w).

The standard prediction concept for the NTU model is stability: A pair (m, w) is a *blocking pair for* X if there exist m' and w' such that $mP_w m'$, $wP_m w'$, $x_{m,w'} > 0$, and $x_{m',w} > 0$; X is *stable* if there are no blocking pairs for X.

We denote by $S(M, W, P, K)$ the set of all stable matchings in $\langle M, W, P, K\rangle$.

2.2. *Matching With Transferable Utility*

The primitives of the model are represented by a four-tuple $\langle M, W, \mathcal{A}, K\rangle$, where M, W, and K are as described above, and $\mathcal{A} = (\alpha_{m,w})$ is an $|M| \times |W|$ matrix of nonnegative real numbers; $\mathcal{A}$ is called a *surplus matrix*, in which $\alpha_{m,w}$ is the surplus jointly generated by a type m man and a type w woman.

Consider the problem

$$\max_{X \in \mathbf{R}_+^{|M|\times|W|}} \sum_{m,w} \alpha_{m,w} x_{m,w} \tag{1}$$

$$\text{such that} \quad \begin{cases} \forall m, & \sum_w x_{m,w} = K_m, \\ \forall w, & \sum_m x_{m,w} = K_w. \end{cases}$$

A matching X is called *optimal* if it is a solution of (1). Optimality is equivalent to stability for the TU model, but not in the NTU model. An optimal matching achieves the maximum total surplus under the population constraints defined by K_m and K_w. It is well known that optimality corresponds to the appropriate notion of stability for the TU model (Shapley and Shubik (1971)). The formal notion of stability requires a discussion of agents' payoffs; for reasons of space, we omit the definition of stability and focus instead on optimal matchings.

2.3. *Extremal Stable Matchings in the NTU Model*

Extremal stable matchings are matchings that are better for one side of the market (say men) and worse for the other side (say women) than any other stable matching. For a type of men (or women), a distribution over the types of women (or men) is preferable over another in the sense of first-order stochastic dominance.

For each $m \in M$, define

$$\mathcal{X}_m = \left\{ x \in \mathbf{R}_+^{|W|} : \sum_{1 \le j \le |W|} x_j = K_m \right\},$$

the set of "distributions" over different types of women that m may match to. Define a partial order $\le_m$ on $\mathcal{X}_m$ by[7]

$$y \le_m x \quad \text{if and only if} \quad \forall w \in W, \sum_{\substack{1 \le j \le |W| \\ w_j R_m w}} y_j \le \sum_{\substack{1 \le j \le |W| \\ w_j R_m w}} x_j.$$

Letting $\mathcal{X}_w = \{x \in \mathbf{R}_+^{|M|} : \sum_{1 \le i \le |M|} x_i = K_w\}$, we define $\le_w$ in an analogous way.

Given an aggregate matching X, let X_m be the row corresponding to type-m men, and let X_w be the column corresponding to type-w women. Baïou and Balinski (2002) (and also Echenique, Lee, and Yenmez (2010)) showed that there are two stable matchings X^M and X^W, such that for any stable matching X,

$$\forall m, \quad X_m^W \le_m X_m \le_m X_m^M,$$
$$\forall w, \quad X_w^M \le_w X_w \le_w X_w^W.$$

We refer to X^M as the *man-optimal* (M-optimal) stable matching and refer to X^W as the *woman-optimal* (W-optimal) stable matching. We also call X^M and X^W *extremal* stable matchings. A matching X is the unique stable matching if $S(M, W, P, K) = \{X\}$; in this case, X coincides with the M- and W-optimal stable matchings.

2.4. *Graph Theoretic Definitions*

To state our main results, we use some basic definitions from graph theory.

A (undirected) *graph* is a pair $G = (V, L)$, where V is a set and L is a subset of $V \times V$. A *path* in G is a sequence $p = \langle v_0, \ldots, v_N \rangle$ such that $(v_n, v_{n+1}) \in L$ for all $n \in \{0, \ldots, N-1\}$. We denote by $v \in p$ that v is a vertex in p. A path $\langle v_0, \ldots, v_N \rangle$ *connects* the vertices v_0 and v_N. A path $\langle v_0, \ldots, v_N \rangle$ is *minimal* if there is no proper subsequence of $\langle v_0, \ldots, v_N \rangle$ that also connects v_0 and v_N.

A *cycle* in G is a path $c = \langle v_0, \ldots, v_N \rangle$ with $v_0 = v_N$. A cycle is *minimal* if for any two vertices v_n and $v_{n'}$ in c, the paths in c from v_n to $v_{n'}$ and from $v_{n'}$ to v_n

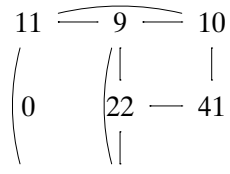
[7]When $K_m = 1$, vectors x and y in $\mathcal{X}_m$ represent probability distributions over the types of women that m may match with. In that case, $y \le_m x$ if and only if the lottery induced by y over W is worse than the lottery induced by x, for any von Neumann–Morgenstern utility function. See also Bogomolnaia and Moulin (2001).

are distinct and minimal. If c and c' are two cycles, and there is a path from a vertex in c to a vertex in c', then we say that c and c' are *connected*.

For an aggregate matching X, we consider the graph defined by letting the vertices be all the nonzero elements of the matrix and the graph defined by letting there be an edge between two vertices when they lie on the same row or column of X. Formally, to each matching X, we associate a graph (V, L) defined as follows. The set of vertices V is $\{(m, w) : m \in M, w \in W \text{ such that } x_{m,w} > 0\}$, and an edge $((m, w), (m', w')) \in L$ is formed for every pair of vertices (m, w) and (m', w') with $m = m'$ or $w = w'$.

Consider the following example, to which we return when discussing our main results. Let matching X be the matrix on the left, with three types of men and women. Its associated graph is depicted on the right:

11	9	10
0	22	41
13	91	0



3. RATIONALIZABILITY

3.1. *Results*

We are now in a position to state the revealed preference problem for matching theory. Given a matching X, we want to understand when there are preferences for M and W such that X is a stable matching or a surplus matrix $\mathcal{A}$ such that X is optimal.

Formally, we say that a matching X has the following qualities:

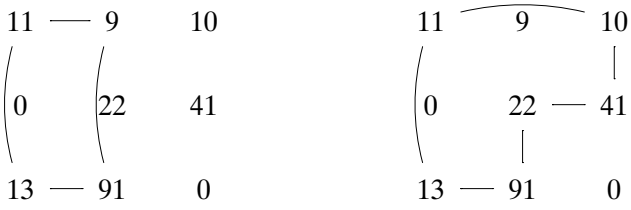
- It is *rationalizable* if there exists a preference profile $P = ((P_m)_{m \in M}, (P_w)_{w \in W})$ such that X is a stable matching in $\langle M, W, P, K \rangle$.
- It is *TU-rationalizable* if there is a surplus matrix $\mathcal{A}$ for which X is the unique solution to problem (1).[8]
- It is *M-optimal* (*W-optimal*) *rationalizable* if there is a profile P for which X is the M-optimal (W-optimal) stable matching in $\langle M, W, P, K \rangle$.

Our main result is a characterization of the rationalizable matchings:

THEOREM 1—Rationalizability: *A matching X is rationalizable if and only if its associated graph does not contain two connected distinct minimal cycles.*

[8]Uniqueness renders the TU-rationalizability problem nontrivial. If we instead required X to be only one of the maximizers of (1), then any matching could be rationalized with a constant surplus ($\alpha_{m,w} = c$ for all m, w). It may be possible to allow for multiple maxima, as long as we impose that the set of maximizers is strictly less than the full set of matchings; we have not explored this possibility here.

As an illustration, consider the matching at the end of Section 2.4. Two minimal cycles in this graph are



The two minimal cycles in the graph are connected; thus, the matching is not rationalizable. In Section 3.2 below, we provide an intuition behind the role of cycles in Theorem 1. The necessity part of the proof of the theorem is in Appendix A and the sufficiency part is in the Supplemental Material (Echenique, Lee, Shum, and Yenmez (2013)).

We now turn to a characterization of TU and extremal rationalizable matchings.

THEOREM 2: *Let X be a matching. The following statements are equivalent*:

(i) *Matching X is TU-rationalizable*.
(ii) *Matching X is M-optimal rationalizable*.
(iii) *Matching X is W-optimal rationalizable*.
(iv) *Matching X is rationalizable as the unique stable matching*.
(v) *The graph associated to X has no minimal cycles*.

This theorem means that the theory of optimal TU matching is strictly more restrictive than NTU stable matching and is observationally equivalent to predicting an extremal, or unique, stable matching. The proof can be found in Appendix B.

3.2. *Intuition*

A crucial distinction between individual and aggregate matchings is the possibility that, say, one type of man may be matched with more than one type of woman and vice versa, resulting in *edges* in the graph corresponding to this matching. To understand their importance for rationalizability, consider a one-to-one individual matching. Obviously, there are no edges here and also the matching is trivially rationalizable: We can set preferences such that each agent's match is his/her most preferred partner, so the observed matching will be stable for these preferences.

In an aggregate matching, however, an edge invalidates these preferences; indeed, for a "vertical" edge (in which two women of the same type match to men of different types), strict preferences imply that, at the least, some women

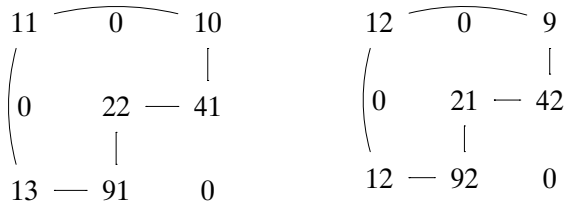
of this type are *not* matched to their most preferred type of men. For this to be stable, it must be that more preferable types of men are "not available" to these women; that is, these men are matched to women whom they find more preferable. Obviously, this imposes restrictions on preferences. In the presence of edges, the rationalizability question boils down, essentially, to the *number* and *configurations* of edges that can be allowed for, such that one can still devise preferences consistent with all the restrictions implied by the edges.

Recall the intuition presented in the Introduction, involving Catherine, Jules, and Jim: by translating the question into the graph defined by the matching, we can get a handle on the problem of the circularity of revealed preferences. Consider the cycle on the left in the example above. This cycle consists of four edges connecting two types of men (call them m_1 and m_3), and two types of women (call them w_1 and w_2). In the cycle, men of type m_1 are matched to women of both types w_1 and w_2. Because of strict preferences, however, m_1 cannot be indifferent between these two types of women. Assuming that women of type w_1 are more preferred (i.e., $w_1 P_{m_1} w_2$), then it must be that for the men of type m_1 who are matched to w_2, the preferable women of type w_1 are not "available" to him; specifically, women of type w_1 who are matched with men of type m_3 must prefer their spouses to men of type m_1 (i.e., $m_3 P_{w_1} m_1$). Obviously, repeating this argument for all four edges in the cycle leads to a very large number of restrictions on the preferences between the types of men and women in the cycle.

In fact, it turns out that there are only *two possible sets* of preference profiles among the four types in the cycle that are consistent with stability. More precisely, preferences within a cycle must be a *flow*; that is, they "point" in one direction. Going back to the left-hand side cycle in the example, this means that if $w_1 P_{m_1} w_2$, then $m_3 P_{w_1} m_1$. To see why this is so, assume to the contrary that $m_1 P_{w_1} m_3$. This would imply that the couples composed of (m_3, w_1) and (m_1, w_2) are unstable, because w_1 from the first couple and m_1 from the second couple would form a blocking pair. Consequently, we also must have $w_2 P_{m_3} w_1$ and $m_1 P_{w_2} m_3$. Graphically, these preferences form a "counterclockwise flow" on the left-hand side matrix of Example 1. Similarly, the "clockwise flow" with preferences satisfying $w_2 P_{m_1} w_1$, $m_3 P_{w_2} m_1$, $w_1 P_{m_3} w_2$, and $m_1 P_{w_1} m_3$ is also consistent with stability.

Theorem 1 says that multiple cycles can coexist in a stable matching only if they are *isolated*; that is, there is no path composed of edges that connect the cycles. Intuitively, this is because when cycles are isolated, the preferences among the types in one cycle do not affect the preferences in another cycle. However, when cycles are connected, then preferences among the types in these cycles are interdependent and, according to Theorem 1, *cannot* be mutually coherent. The issue is that preferences along a path that connects two cycles must also be a flow, but the flow must point away from each cycle. It is not possible for a flow to point away from both of the connected cycles.

Now, to illustrate the ideas behind Theorem 2, we present a slightly modified example. Consider the aggregate matching on the left:



The matching exhibits a unique cycle, so it is (NTU–) rationalizable. Assume, contradicting Theorem 2, that it is rationalizable as an M-optimal stable matching. Label the agents according to the row or column number. By the argument we made above, the rationalizing preferences must define a flow; say that $w_1 P_{m_1} w_3$, $m_3 P_{w_1} m_1$, and so on. Then we can create a stable matching that is better for men and worse for women: Shift one type-m_1 man from a type-w_3 woman to match a type-w_1 woman, which we take from a type-m_3 man. Note that this man is better off and the woman is worse off. Such a shift would leave a type-w_3 woman and a type-m_3 man unmatched, but because there is a cycle, we can complete a sequence of shifts that defines a new matching, one in which some men are better off and some women are worse off, while the rest of the agents are indifferent. This is the matching in the matrix on the right.

A similar argument underlies the TU-rationalizability result; it turns out that when a cycle is present, it is always possible to increase the total surplus of agents by "shifting" agents within the cycle, as we did above.

3.3. *Discussion*

Theorems 1 and 2 provide a complete picture of the empirical content of models of stable matching. We conclude by discussing some implications and qualifications of our results.

Throughout, we restrict attention to the rationalizability question when only matching data are available to the researcher; particularly, no data on transfers are available. This appears reasonable, especially as a prominent application of the TU matching model in empirical work has been to the marriage market, in which no explicit transfers are observed. In applications to labor markets, however, transfers may be observable.

We reiterate here that our results are based on the assumption that agents of the same type are identical: they have identical preferences and they are perceived as identical by all other agents. This assumption is standard in most exercises on revealed preferences, but it is problematic.[9]

[9]Most empirical studies of revealed preference in consumption, from Famulari (1995) to Blundell, Browning, and Crawford (2003), make the same assumption.

It is possible that an observed matching may fail to be rationalizable because we have imposed a too rigid structure on individual preferences. In actual empirical implementations of our test, one would need to introduce some additional flexibility (possibly in the form of measurement errors, as in Varian (1985) or Echenique, Lee, and Shum (2011) in the context of consumption behavior). Finally, allowing for too much heterogeneity in individual preferences renders the theory nontestable, so it is also possible to interpret the strong conditions we obtain in our paper as a negative result.

Our results imply that the matching model with transfers is nested in the model without transfers and that a stable matching with transfers is observationally equivalent to extremal, and unique, stable matchings without transfers.

Such a relation between these models is striking, especially in light of the fact that the TU and extremal stable matching models are the two most widely used models in applied work on matching markets. Particularly, most econometric studies on matchings assume that there are transfers, motivated by the view in Becker (1973) that transfers in matching may be implicit and nonpecuniary. On the other hand, most applications of matching in market design seek to implement an extremal stable matching, using the algorithm of Gale and Shapley (1962). When preferences are unknown, there are no reasons to a priori expect one theory to be more relevant than the other: the theories should be compared empirically. *Our results imply that when only observing an aggregate matching, the matching theory with transfers is nested in the matching theory without transfers, and the predictions of the Gale–Shapley algorithm are equivalent to Becker's model of marriage with transfers.*

APPENDIX A: PROOF OF THEOREM 1

A.1. *Proof of Necessity*

Let (V, L) be the graph defined from a matching X.

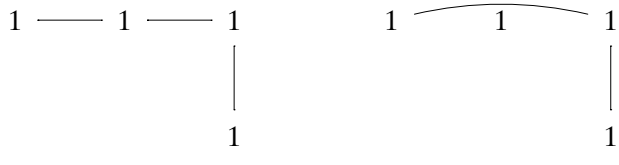
We start with two simple facts about minimal cycles and paths:

(a) If $c = \langle v_0, \ldots, v_N \rangle$ is a minimal cycle, then no vertex other than $v_0 = v_N$ appears twice in c.

(b) Let $\{(m, w)_n : n = 0, \ldots, N\}$ be a minimal path with $N \geq 2$. Then for any $n \in \{0, \ldots, N-2\}$,

$$(m_n = m_{n+1} \Rightarrow w_{n+1} = w_{n+2}) \quad \text{and} \quad (w_n = w_{n+1} \Rightarrow m_{n+1} = m_{n+2});$$

that is, any two subsequent edges in a minimal path must be at right angles:



The path on the left is not minimal; the path on the right is.

An *orientation* of (V, L) is a mapping $d: L \to \{0, 1\}$. We often write $d((m, w), (m, w'))$ as $d_{m,w,w'}$ and $d((m, w), (m', w))$ as $d_{w,m,m'}$. A preference profile P defines an orientation d by setting $d_{m,w,w'} = 1$ if and only if $wP_m w'$, and $d_{w,m,m'} = 1$ if and only if $mP_w m'$.

Let d be an orientation defined from a preference profile. Then X is stable if and only if, for all (m_1, w_1) and (m_2, w_2), if $x_{1,1} > 0$ and $x_{2,2} > 0$, then

$$d_{m_1,w_2,w_1} d_{w_2,m_1,m_2} = 0 \quad \text{and} \quad d_{m_2,w_1,w_2} d_{w_1,m_2,m_1} = 0. \tag{2}$$

We say that the pair $((m_1, w_1), (m_2, w_2))$ is an *antiedge* if $x_{1,1} > 0$ and $x_{2,2} > 0$ for $m_1 \neq m_2$ and $w_1 \neq w_2$.

A path $\{(m, w)_n : n = 0, \ldots, N\}$ is a *flow* for d if either $d((m, w)_n, (m, w)_{n+1}) = 1$ for all $n \in \{0, \ldots, N-1\}$ or $d((m, w)_n, (m, w)_{n+1}) = 0$ for all $n \in \{0, \ldots, N-1\}$. If the second statement is true, we call the path a *forward flow*.

Fix an orientation d derived from the preferences that rationalize X.

LEMMA 1: *Let $p = \langle (m, w)_n : n = 0, \ldots, N \rangle$ be a minimal path. If $d((m, w)_0, (m, w)_1) = 0$ or $d((m, w)_{N-1}, (m, w)_N) = 1$, then p is a flow for d.*

PROOF: Because subsequent edges in a minimal cycle form right angles, for any $n \in \{1, \ldots, N-1\}$, the pair of vertices $(m, w)_{n-1}$ and $(m, w)_{n+1}$ form an antiedge: we have $x_{(m,w)_{n-1}} > 0$, $x_{(m,w)_{n+1}} > 0$, $m_{n-1} \neq m_{n+1}$, and $w_{n-1} \neq w_{n+1}$. Further, $(m, w)_n$ has one element in common with $(m, w)_{n-1}$ and the other in common with $(m, w)_{n+1}$. Therefore, if $d((m, w)_{n-1}, (m, w)_n) = 0$ or, equivalently, $d((m, w)_n, (m, w)_{n-1}) = 1$, then we have $d((m, w)_n, (m, w)_{n+1}) = 0$ by equation (2).

The argument in the previous paragraph shows that the existence of some n' with $d((m, w)_{n'-1}, (m, w)_{n'}) = 0$ implies $d((m, w)_{n-1}, (m, w)_n) = 0$ for all $n \geq n'$. So if $d((m, w)_0, (m, w)_1) = 0$, then $d((m, w)_n, (m, w)_{n+1}) = 0$ for all $n \in \{1, \ldots, N-1\}$, and if $d((m, w)_{N-1}, (m, w)_N) = 1$, then $d((m, w)_n, (m, w)_{n+1}) = 1$ for all $n \in \{0, \ldots, N-1\}$. In either way, p is a flow. *Q.E.D.*

We obtain the following lemma from Lemma 1.

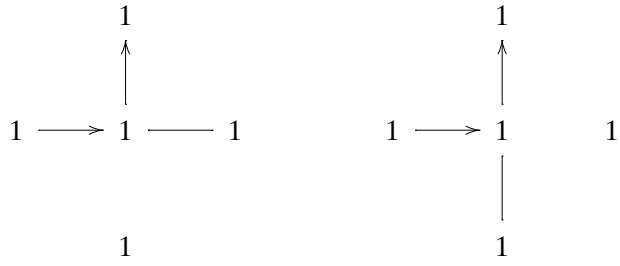
LEMMA 2: *If $c = \langle (m, w)_n \rangle$ is a minimal cycle, then c is a flow for d.*

Let $p = \langle (m, w)_n \rangle$ be a path and let $(m, w) \notin p$. We say that a path $\bar{p} = \langle (\bar{m}, \bar{w})_n : n = 0, \ldots, \bar{N} \rangle$ *connects* p and (m, w) if $(\bar{m}, \bar{w})_0 \in p$ and $(\bar{m}, \bar{w})_{\bar{N}} = (m, w)$.

LEMMA 3: *Let $c = \langle (m, w)_n : n = 0, \ldots, N \rangle$ be a minimal cycle, and let $p = \langle (\bar{m}, \bar{w})_n : n = 0, \ldots, \bar{N} \rangle$ be a minimal path connecting c and $(\bar{m}, \bar{w}) \notin c$. Then $\langle (\bar{m}, \bar{w})_n : n = 1, \ldots, \bar{N} \rangle$ is a forward flow.*

PROOF: By Lemma 2, c is a flow for d. Suppose that c is a forward flow. If c is not a forward flow, we can re-index $(m, w)_k$ by $(m, w)_{N-k}$, thereby making c be a forward flow. For any positive integer r, we define $(m, w)_r$ to be $(m, w)_k$, where k is the remainder when we divide r by N (e.g., $(m, w)_{2N+1} = (m, w)_1$). In doing so, we can index the cycle by any positive integer.

To prove Lemma 3, we need to deal with two cases. Let $(m, w)_{n^*} = (\bar{m}, \bar{w})_0$. By definition of a cycle, $(\bar{m}, \bar{w})_0$ shares either m or w with $(m, w)_{n^*-1}$. Suppose, without loss of generality, that they share m, so $\bar{m}_0 = m_{n^*-1}$. The two cases in question are represented below, where the center vertex is $(\bar{m}, \bar{w})_0$. Case 1 on the left has $(\bar{m}, \bar{w})_1$ also sharing m with $(\bar{m}, \bar{w})_0$, while Case 2 has $(\bar{m}, \bar{w})_0$ sharing w with $(\bar{m}, \bar{w})_1$.



Case 1. Suppose that $\bar{m}_1 = \bar{m}_0 = m_{n^*-1}$. Consider the minimal path

$$p' = \left\langle (m, w)_{n^*-1}, (\bar{m}, \bar{w})_1, \ldots, (\bar{m}, \bar{w})_{\bar{N}} \right\rangle.$$

Since $m_{n^*-2} \neq \bar{m}_1$, the path

$$\hat{p} = \left\langle (m, w)_{n^*-2}, (m, w)_{n^*-1}, (m, w)_1 \right\rangle$$

is a minimal path from $(m, w)_{n^*-2}$ to $(m, w)_1$. We have that $d((m, w)_{n^*-2}, (m, w)_{n^*-1}) = 0$, as c is a forward flow. It follows by Lemma 1 that $d((m, w)_{n^*-1}, (\bar{m}, \bar{w})_1) = 0$ and thus $\hat{p}$ is also a forward flow. Then, by Lemma 1 again, p' is a forward flow; in particular, $d((\bar{m}, \bar{w})_n, (\bar{m}, \bar{w})_{n+1}) = 0$ for $n \in \{1, \ldots, \bar{N} - 1\}$.

Case 2. Suppose that $\bar{m}_1 \neq \bar{m}_0 = m_{n^*-1}$. Then the path

$$\left\langle (m, w)_{n^*-1}, (\bar{m}, \bar{w})_0, (\bar{m}, \bar{w})_1 \right\rangle$$

is a minimal path connecting $(m, w)_{n^*-1}$ and $(\bar{m}, \bar{w})_1$. We have that $d((m, w)_{n^*-2}, (m, w)_{n^*-1}) = 0$, as c is a forward flow. By an application of Lemma 1, analogous to the one in Case 1, we obtain that p is a forward flow.

Regardless of whether we are in Case 1 or Case 2, the path $\langle (\bar{m}, \bar{w})_n : n = 1, \ldots, \bar{N} \rangle$ is a forward flow. *Q.E.D.*

Lemma 4 finishes the proof of the necessity direction.

LEMMA 4: *There are no two connected distinct minimal cycles.*

PROOF: Suppose, for contradiction, that there are two minimal cycles c_1 and c_2, and a path $p = \langle (m,w)_n : n = 0, \ldots, N \rangle$ connecting $(m,w)_0 \in c_1$ with $(m,w)_N \in c_2$. We can suppose, without loss of generality, that p is minimal. We can also suppose that $N \geq 3$, because if $N < 3$, we can add $(m',w') \in c_1$ with $((m',w'),(m,w)_0) \in L$, and $(m'',w'') \in c_2$ with $((m'',w''),(m,w)_N) \in L$ to p; the corresponding path will also be a minimal path connecting c_1 and c_2.

By applying Lemma 3 to c_1 and p, we obtain that $\langle (m,w)_n : n = 1, \ldots, N \rangle$ is a forward flow. By applying Lemma 3 to c_2 and p, we also find that $\langle (m,w)_{N-k} : k = 1, \ldots, N \rangle$ is a forward flow. The first statement implies that $d((m,w)_1,(m,w)_2) = 0$, and the second implies that $d((m,w)_1,(m,w)_2) = 1$, which contradict each other. *Q.E.D.*

A.2. *Proof of Sufficiency: An Illustration*

The proof of the sufficiency direction of Theorem 1 is constructive; it works by using an algorithm to construct a rationalizing preference profile. Since it is rather tedious, we present an example that illustrates how the algorithm works. For a detailed proof, see the Supplemental Material.

EXAMPLE 1—Constructing Rationalizing Preferences: Consider the matching.

$$X = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

The algorithm first identifies the unique minimal cycle (if it exists), and then finds minimal paths connecting the cycle and vertices not in the cycle by searching over the graph (V, L). In our example, there is a minimal cycle $\{(m_1,w_1),(m_4,w_1),(m_4,w_3),(m_1,w_3)\}$. From the cycle, we denote the set of types of men and the set of types of women in the cycle by $\bar{M}_1$ and $\bar{W}_1$, respectively:

$$\bar{M}_1 = \{m_1, m_4\} \quad \text{and} \quad \bar{W}_1 = \{w_1, w_3\}.$$

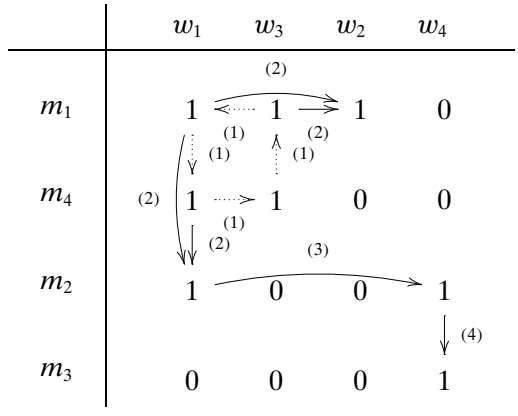
Subsequently, we define $\bar{M}_k$ as the set of types of men that are not in $\bigcup_{k'=1}^{k-1} \bar{M}_{k'}$ and that match to types of women in $\bar{W}_{k-1}$. We similarly define $\bar{W}_k$; that is,

$$\bar{M}_2 = \{m_2\}, \quad \bar{W}_2 = \{w_2\},$$
$$\bar{M}_3 = \emptyset, \quad \bar{W}_3 = \{w_4\},$$

and

$$\bar{M}_4 = \{m_3\}, \quad \bar{W}_4 = \emptyset.$$

We orient (V, L) such that the minimal cycle is a flow, and each minimal path connecting the cycle and a vertex not in the cycle is a forward flow. Accordingly, we obtain the following graph.



All orientations labeled (1) show that the minimal cycle is a flow. The orientations denoted (2), (3), and (4) are determined as we sequentially specify minimal paths connecting the cycle and the vertices not in the cycle, which are all forward flows.

Given the constructed orientation d, we define two collections of partial orders: $(\tilde{P}_m : m \in M)$ and $(\tilde{P}_w : w \in W)$: $w\tilde{P}_m w'$ when $d_{m,w,w'} = 1$, and $m\tilde{P}_w m'$ when $d_{w,m,m'} = 1$. Then we extend $\tilde{P}_m$ by including $w\tilde{P}_m w'$ if $x_{m,w} > 0$ and $x_{m,w'} = 0$. This extended preference is a well defined strict partial order. Thus, we can extend the preference of man type m further to be a complete strict order on W. We similarly extend preferences of all other types of men and types of women. These preferences rationalize the matching X.

APPENDIX B: PROOF OF THEOREM 2

We proceed by proving first that rationalizability as either M- or W-optimal stable matching (i.e., extremal rationalizability) implies the absence of cycles. Second, we prove that the absence of cycles implies rationalizability as a unique stable matching. Since a unique stable matching is trivially both M- and W-optimal, the condition also implies rationalizability as extremal stable matchings.

We omit the proof that TU-rationalizability is equivalent to the absence of cycles. The proof can be found in Echenique, Lee, and Yenmez (2010).

B.1. *Proof That Extremal Rationalizability Implies the Absence of Cycles*

Let X be a matching that is extremal rationalizable. There exists a preference profile P such that X is M-optimal or W-optimal stable matching in $\langle M, W, P, K\rangle$.

Suppose, for contradiction, that the graph (V, L) associated to X has a minimal cycle $c = \langle v_0, \ldots, v_N\rangle$. As before, we denote m_n for the type of the men in v_n and denote w_n for the type of the women in v_n, respectively.

As we do in the beginning of Section A.1, we construct an orientation d from the preference profile P that rationalizes X as an extremal matching. According to Lemmas 1 and 2, we can index c such that the path $\langle v_n\rangle_{n=0}^{N-1}$ is a flow for d such that $d_{n,n,n+1} = 0$ for all $n = 0, 1, \ldots, N-1$; that is, if an edge (v_n, v_{n+1}) is vertical (i.e., $w_n = w_{n+1}$), we have $d_{w_n,m_n,m_{n+1}} = 0$, and when the edge is horizontal (i.e., $m_n = m_{n+1}$), we have $d_{m_n,w_n,w_{n+1}} = 0$.

We introduce two partial orders on matchings. For two matchings X and Y,

$$X \leq_M Y \quad \text{if, for all } m,\ X_m \leq_m Y_m,$$
$$X \leq_W Y \quad \text{if, for all } w,\ X_w \leq_w Y_w.$$

In the following proof, we show that we can make the types of men (women) weakly better (worse) off by "rematching" agents whose matches are involved in the cycle c while preserving stability. We can also make types of women (men) weakly better (worse) off with a similar rematching. Therefore, X is neither M-optimal nor W-optimal stable matching.

We capture "rematching" using a matrix of differences in matches: Let $\mathcal{E}$ be the set of all $|M| \times |W|$ matrices E such that for all i and j.

(i) $e_{i,j} = 0$ if and only if (i, j) is *not* in the cycle c,
(ii) $\sum_{h=1}^{|W|} e_{i,h} = 0$ and $\sum_{l=1}^{|M|} e_{l,j} = 0$,
(iii) $|e_{i,j}| \leq x_{i,j}$.

CLAIM 1: *For all $E \in \mathcal{E}$, $X + E$ and $X - E$ are stable in $\langle M, W, P, K\rangle$, and either*

$$X - E \leq_M X \leq_M X + E \quad \textit{and} \quad X - E \geq_W X \geq_W X + E$$

or

$$X + E \leq_M X \leq_M X - E \quad \textit{and} \quad X + E \geq_W X \geq_W X - E.$$

PROOF: For any $E \in \mathcal{E}$, $X + E$ is a well defined matching: by property (ii), the row and column sum of $X + E$ respect the feasibility constraints; by property (iii), the entries of $X + E$ are nonnegative; by property (i), the matrix $X + E$ is also a stable matching, as

$$x_{i,j} + e_{i,j} > 0 \quad \Rightarrow \quad x_{i,j} > 0;$$

that is, if there were a blocking pair of type m_i and type w_j under $X + E$, it would also be a blocking pair under X. Since $E \in \mathcal{E} \Rightarrow -E \in \mathcal{E}$, $X - E$ is also well defined and stable.

As a consequence of properties (i) and (ii), e_n (i.e., e_{m_n,w_n}) alternates in sign as n increases: if $e_n > 0$, then e_{n+1} (i.e., $e_{m_{n+1},w_{n+1}}$) is less than 0. This implies that one of the following two cases has to hold.

(a) For all n, if $m_n = m_{n+1} = m$, then $e_{m,w_n} < 0$ and $e_{m,w_{n+1}} > 0$, and if $w_n = w_{n+1} = w$, then $e_{m_n,w} > 0$ and $e_{m_{n+1},w} < 0$.

(b) For all n, if $m_n = m_{n+1} = m$, then $e_{m,w_n} > 0$ and $e_{m,w_{n+1}} < 0$, and if $w_n = w_{n+1} = w$, then $e_{m_n,w} < 0$ and $e_{m_{n+1},w} > 0$.

We proceed by assuming that we are in case (a) and we prove that $X - E \leq_M X \leq_M X + E$. It will become clear that if we were in case (b), we would establish that $X + E \leq_M X \leq_M X - E$.

Fix $m \in M$. By definition of minimal cycle, there is at most one n such that $v_n, v_{n+1} \in c$ and $m_n = m_{n+1} = m$. If no such n exists, $E_m = 0$ by property (i) of $\mathcal{E}$ and, thus, trivially $(X - E)_m \leq_m X_m \leq_m (X + E)_m$. On the other hand, if there is n such that $v_n, v_{n+1} \in c$ and $m_n = m_{n+1} = m$, then (v_n, v_{n+1}) is horizontal and (v_{n+1}, v_{n+2}) is vertical. From the orientation d, we have $d_{m,w_n,w_{n+1}} = 0$, implying that $w_{n+1} P_m w_n$. In E_m, only e_n and e_{n+1} are nonzero, and $e_{n+1} = -e_n > 0$ as we consider the case (a). By definition of $\leq_m$, $w_{n+1} P_m w_n$ implies that $(X - E)_m \leq_m X_m \leq_m (X + E)_m$.

Since the type m was arbitrary, we obtain $X - E \leq_M X \leq_M X + E$. By Theorem 5 in Baïou and Balinski (2002) (and also by Theorem 2 in Echenique, Lee, and Yenmez (2010)), this also implies that $X - E \geq_W X \geq_W X + E$. Last, $X \neq X + E$ and $X \neq X - E$ by property (i) of $\mathcal{E}$, implying that X is neither M-optimal nor W-optimal stable matching. *Q.E.D.*

B.2. *Proof That the Absence of Cycles Implies Unique Rationalizability*

We prove that if the graph (V, L) associated to X has no cycles, then there is a preference profile P such that $\langle M, W, P, K \rangle$ has X as its unique stable matching. The matching X is, therefore, both M- and W-optimal stable matchings.

We consider a particular set of preferences. Let $U = (u_{m,w}) \in \mathbf{R}_+^{|M| \times |W|}$ in which $u_{m,w} \neq u_{m',w'}$ for all $(m, w) \neq (m', w')$. For each m and w, a man of type m and a woman of type w both receive an equal utility $u_{m,w}$ by being matched to each other. We consider the preference profile induced by U, which we denote by P_U and call *a perfectly correlated preference profile*.

LEMMA 5: *If a preference profile is perfectly correlated, there exists a unique stable matching.*

PROOF: Suppose, for contradiction, that X and Y are two distinct stable matchings. Let $\mathcal{U}$ be the set of numbers $u_{m,w}$ for m and w such that $x_{m,w} \neq y_{m,w}$.

Let (m^*, w^*) be such that $u_{m^*,w^*} \in \mathcal{U}$ and $u_{m^*,w^*} \geq u$ for all $u \in \mathcal{U}$. Suppose, without loss of generality, that $x_{m^*,w^*} < y_{m^*,w^*}$. Note that

$$\sum_{m:mP_{w^*}m^*} x_{m,w^*} = \sum_{m:mP_{w^*}m^*} y_{m,w^*}, \qquad \sum_{w:wP_{m^*}w^*} x_{m^*,w} = \sum_{w:wP_{m^*}w^*} y_{m^*,w},$$

because $mP_{w^*}m^* \Rightarrow u_{m,w^*} > u_{m^*,w^*} \Rightarrow x_{m,w^*} = y_{m,w^*}$, and similarly for the second equality.

Since $x_{m^*,w^*} < y_{m^*,w^*}$, there exist m and w such that $x_{m,w^*} > 0$, $x_{m^*,w} > 0$, $w^*P_{m^*}w$, and $m^*P_{w^*}m$. Thus, (m^*, w^*) is a blocking pair of X, contradicting the stability of X. *Q.E.D.*

Suppose that the graph (V, L) associated to X has no minimal cycles, so it has no cycles. Using the absence of cycles, we assign cardinal utilities $U = (u_{m,w})$, such that X is a stable matching for $\langle M, W, P_U, K\rangle$. Lemma 5 then guarantees that X is the unique stable matching. We first consider the case when all nodes in V are connected and later generalize to the case in which there are multiple connected components of (V, L).

Choose a vertex v_0 in V. Since (V, L) contains no cycles, for each $v \in V$ there is a unique minimal path connecting v_0 to v in (V, L). Let $\eta(v)$ be the length of the minimal path connecting v_0 to v. For $v \in V$ (i.e., $x_v > 0$), assign $u_v = (1 + \eta(v)) + \varepsilon_v$, and for all other (m, w) with $x_{m,w} = 0$, assign $u_{m,w} = \varepsilon_{m,w}$. All ε_v and $\varepsilon_{m,w}$ are positive, and distinct real numbers; we assume all $\varepsilon_{m,w}$ are small enough that if $\eta(v) > \eta(v')$, then $u_v > u_{v'}$.[10] We suppose that both a type-m man and a type-w woman receive the same utility $u_{m,w}$ by being matched to each other.

To show that X is a (unique) stable matching in $\langle M, W, P_U, K\rangle$, suppose, for contradiction, that a pair (m_i, w_j) blocks X: there exist $m_{i'}$ and $w_{j'}$ such that $x_{i,j'} > 0$, $x_{i',j} > 0$, $u_{i,j} > u_{i,j'}$, and $u_{i,j} > u_{i',j}$. Since $x_{i,j'} > 0$ and $x_{i',j} > 0$, they are nodes in V, and thus $u_{i,j'} > 1$ and $u_{i',j} > 1$. Then, by definition of U, we have $u_{i,j} > \max\{u_{i',j}, u_{i,j'}\} > 1$, which implies $x_{i,j} > 0$. In all, $\langle (m_{i'}, w_j), (m_i, w_j), (m_i, w_{j'})\rangle$ is a path.

There are unique paths from v_0 to each $(m_{i'}, w_j)$, (m_i, w_j), and $(m_i, w_{j'})$. Moreover, $v_0 \neq (m_i, w_j)$ since $u_{i,j} > u_{i',j}$. Consider the minimal path $p = \langle v_0, \ldots, v_N\rangle$ connecting v_0 to (m_i, w_j); that is, $v_N = (m_i, w_j)$. If this path does not include $(m_{i'}, w_j)$, then the minimal path connecting v_0 to $(m_{i'}, w_j)$ is $\langle v_0, \ldots, v_N, (m_{i'}, w_j)\rangle$ since there exists no cycle. Therefore, $u_{i',j} > u_{i,j}$ which is a contradiction. We get a similar contradiction when p does not include $(m_i, w_{j'})$. But since p is minimal, p cannot contain both $(m_i, w_{j'})$ and $(m_{i'}, w_j)$. Consequently, (m_i, w_j) cannot be a blocking pair.

When (V, L) has multiple components $\{(V_1, L_1), \ldots, (V_N, L_N)\}$, we can partition M and W as $(M_1, \ldots, M_N)$ and $(W_1, \ldots, W_N)$ such that for all $v \in V_n$, we

[10] We use ε_v and $\varepsilon_{m,w}$ only to ensure strict preferences.

have $m_v \in M_n$ and $w_v \in W_n$. For each (V_n, L_n) with associated sets M_n and W_n, we assign utilities $(u_{m,w})_{(m,n)\in M_n \times W_n}$ similar to the single component case. For other $(m, w) \in M_n \times W_l$ with $n \neq l$, we assign $u_{m,w} = \varepsilon_{m,w}$, which are all small and positive real numbers, and $\varepsilon_{m,w} \neq \varepsilon_{m',w'}$ when $(m, w) \neq (m', w')$.

Suppose a type-m man and a type-w woman are not matched under X. If there is n such that $(m, w) \in M_n \times W_n$, then (m, w) is not a blocking pair by the proof above for the case of a single connected component. If $(m, w) \in M_n \times W_l$ with $n \neq l$, then, by the construction of $u_{m,w}$, $w' P_m w$ for every w' with $x_{m,w'} > 0$. Thus, (m, w) is again not a blocking pair, and X is a stable matching and is the unique stable matching by Lemma 5.

REFERENCES

ABDULKADIROĞLU, A., P. PATHAK, A. ROTH, AND T. SÖNMEZ (2005): "The Boston Public School Match," *American Economic Review*, 95 (2), 368–371. [155]

BAÏOU, M., AND M. BALINSKI (2002): "Erratum: The Stable Allocation (or Ordinal Transportation) Problem," *Mathematics of Operations Research*, 27 (4), 662–680. [157,168]

BECKER, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81 (4), 813–846. [154,162]

BLUNDELL, R., M. BROWNING, AND I. CRAWFORD (2003): "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71 (1), 205–240. [161]

BOGOMOLNAIA, A., AND H. MOULIN (2001): "A New Solution to the Random Assignment Problem," *Journal of Economic Theory*, 100 (2), 295–328. [157]

CHAMBERS, C. P., AND F. ECHENIQUE (2009): "The Core Matchings of Markets With Transfers," SS Working Paper 1298, Caltech. [155]

CHOO, E., AND A. SIOW (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114 (1), 175–201. [154,155]

ECHENIQUE, F. (2008): "What Matchings Can Be Stable? The Testable Implications of Matching Theory," *Mathematics of Operations Research*, 33 (3), 757–768. [155]

ECHENIQUE, F., S. LEE, AND M. SHUM (2010): "Aggregate Matchings," SS Working Paper 1316, Caltech. [153]

——— (2011): "The Money Pump as a Measure of Revealed Preference Violations," *Journal of Political Economy*, 119, 1201–1223. [162]

ECHENIQUE, F., S. LEE, M. SHUM, AND M. B. YENMEZ (2013): "Supplement to 'The Revealed Preference Theory of Stable and Extremal Stable Matchings'," *Econometrica Supplemental Material*, 81, http://www.econometricsociety.org/ecta/Supmat/10011_Proofs.pdf. [159]

ECHENIQUE, F., S. LEE, AND M. B. YENMEZ (2010): "Existence and Testable Implications of Extreme Stable Matchings," SS Working Paper 1337, Caltech. [153,155,157,166,168]

FAMULARI, M. (1995): "A Household-Based, Nonparametric Test of Demand Theory," *Review of Economics and Statistics*, 77 (2), 372–382. [161]

FOX, J. T. (2008): "Estimating Matching Games With Transfers," WP 14382, NBER. [154,155]

GALE, D., AND L. S. SHAPLEY (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69 (1), 9–15. [154,162]

GALICHON, A., AND B. SALANIE (2009): "Matching With Trade-Offs: Revealed Preferences Over Competing Characteristics," Report, Ecole Polytechnique. [155]

KESTEN, O., AND U. ÜNVER (2009): "A Theory of School Choice Lotteries," Report, Boston College and Carnegie Mellon University. [155]

ROTH, A. E. (2002): "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, 70 (4), 1341–1378. [154]

——— (2008): "Deferred Acceptance Algorithms: History, Theory, Practice and Open Questions," *International Journal of Game Theory*, 36 (3), 537–569. [154]

SHAPLEY, L., AND M. SHUBIK (1971): "The Assignment Game I: The Core," *International Journal of Game Theory*, 1 (1), 111–130. [154,156]

SÖNMEZ, T., AND U. ÜNVER (2011): "Matching, Allocation, and Exchange of Discrete Resources," in *Handbook of Social Economics*, Vol. 1A, ed. by J. Benhabib, A. Bisin, and M. Jackson. Amsterdam, The Netherlands: North-Holland, 781–852. [154]

SØRENSEN, M. (2007): "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital," *Journal of Finance*, 62 (6), 2725–2762. [154]

VARIAN, H. R. (1985): "Non-Parametric Analysis of Optimizing Behavior With Measurement Error," *Journal of Econometrics*, 30, 445–458. [162]

Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, U.S.A.; fede@caltech.edu,

Dept. of Economics, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.; sangmok@sas.upenn.edu,

Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, U.S.A.; mshum@caltech.edu,

and

Tepper School of Business, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, U.S.A.; byenmez@andrew.cmu.edu.

Manuscript received May, 2011; final revision received March, 2012.